IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning Techniques For Leukemia Classification: A Literature Review

Simran S Shende [1], Dr. A.P. Thakare[2], Prof. P.P. Likhitkar[3]

Department of Electronics and Telecommunication Engineering

[1]Dr. Rajendra Gode Institute of Technology and Research Amravati.

[2] Professor and Head of Department, Sipna College of Engineering and Technology, Amravati

[3] Assistant Professor Dr. Rajendra Gode Institute of Technology and Research Amravati.

Abstract

Leukemia, also known as blood cancer is a dangerous disease that arises in the bone marrow and overproduces malformed white blood cells. These cells make blood unhealthy and weaken body's ability to fight with infection. Leukemia is categorized based on speed of progression [Acute and Chronic] and the types of white blood cell affected [lymphocytic or myeloid]. Due to high risk factor, it requires accurate and timely diagnosis for effective treatment. Current advancements in Machine learning (ML) have revealed promising diagnostic results. This study aims to provides a complete overview of different Machine learning and Deep learning Techniques applied for the diagnosis of leukemia, highlighting the parameters like accuracy and number of features extracted, using different segmentation techniques.The paper explores the application of various Machine Learning (ML) algorithms, with Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), and Image Processing techniques, demonstrating outstanding results in a specific area. These algorithms have shown outstanding success in tasks like image classification, object detection, and medical image analysis. Our review highlights the capability of ML in enhancing leukemia diagnosis accuracy, reducing diagnosis time, and improving patient outcomes.

Keywords: Leukemia, image processing, Machine learning, SVM.

Introduction:

Leukaemia is responsible for almost 2.5% of all new cancer occurrence and 3.1% of cancer-related mortality [1]. According to FAB [French-America-British] the main types of leukaemia include Acute Lymphoblastic Leukaemia (ALL), Acute Myeloid Leukaemia (AML), Chronic Lymphocytic Leukaemia (CLL), and Chronic Myeloid Leukaemia (CML) [2], while leukemia can arise in both adults and children. Leukemia diagnostic tests, which usually consist of blood tests to measure blood cell counts RBC [Red blood cell], WBC [White blood cell] and platelets. To detect abnormal cells various segmentation methods are used to analyze textural, morphological and statistical parameters of blood cell also, bone marrow investigations is used to determine whether leukemia cells are existing in the bone marrow, and other imaging procedures like CT scan and X-rays, to investigate early detection and proper utilization of technology.

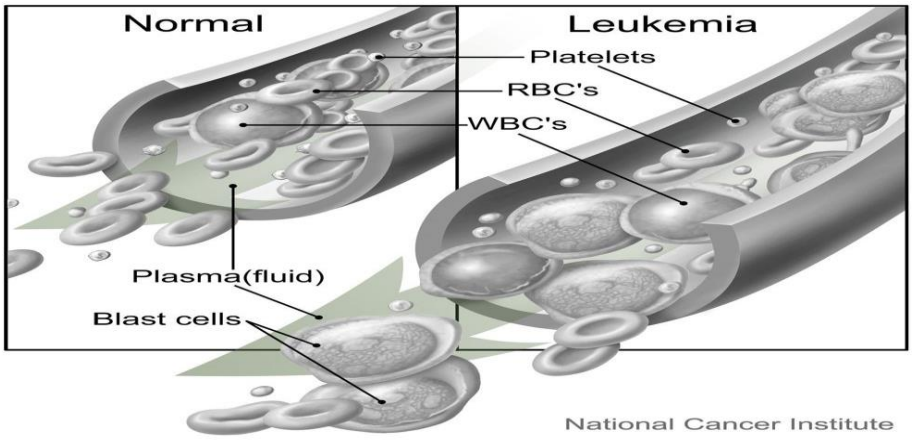

Fig.1 Normal and Abnormal blood cells

[https://www.roswellpark.org/sites/default/files/styles/max_2600x2600/public/leukemia-diagram-nci.jpg?itok=_qCIgegh]

Literature Review:

As the manual methods of detecting blood cancer are time consuming and tedious. it requires skilled operators for accurate detection of disease. On the other hand various AI techniques gaining advancement in identifying types of cancer. To detect Leukemia several parameters of white blood cell have to examine. lymphocytes are irregular in shapes and have compact nucleus with definite and sharp boundaries whereas lymphoblasts are irregular in shape consist of cavities in cytoplasm [3]. Putzu et.al identified 245 out of 267 total leucocytes accurately, yielding an accuracy of 92% across 33 images captured under consistent camera and lighting conditions using SVM achieving an accuracy of 93% image processing technique and extracted different features like cell shape, color and texture with high level of accuracy [4]

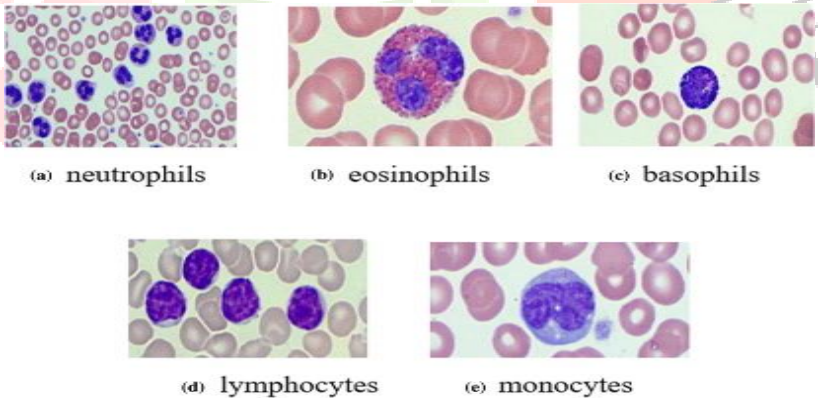

Fig 2: Classification of WBC (Courtesy: https://medschool.co/tests/blood-film)

While in paper [5] mainly focuses on detecting and comparing two types of leukemia namely, ALL and AML with the Customized Convolution neural network model (based on deep learning) of Tversky loss function has been used to study multiple layers of images and investigated parameters like precision, sensitivity, kappa, fscore with accuracy of 99%.

Fatma Talaat and Samah Gamel proposed an image processing approach integrated with fuzzy optimization to build up the accuracy of leukemia detection. This methodology involved an optimized CNN classification for extracting features from images and differentiate between normal and abnormal leukemia cases. With C-NMC dataset researcher used fuzzy logic to analyses hyperparameters to get best result [6]

The research [7] introduced three filtration techniques which enhance the accuracy of Acute Lymphoblastic Leukemia detection and sixteen Morphological features were extracted through image processing, focusing on parameters such as form factor, eccentricity, elongation, and cell dissimilarity. To perform classification two classifier has been used namely Artificial Neural Network and Support Vector Machine with overall accuracy of 97.52%. [7]

The study [8] introduced a new approach to detect leukemia by combining methods like Image Processing, Machine learning and Transfer learning model conducts an in-depth evaluation of machine models for early leukemia diagnosis. With diverse Kaggle dataset utilization for binary and multiclass classification the Inception-Res-Net model demonstrated outstanding result.[8]

Bodzas et.al proposed conventional digital image processing and ML algorithm automatically identify cancer. Moreover, sixteen robust characteristics of microscopic images were found. For further classification purpose SVM and ANN classifier is used with overall accuracy of 97.52% [9]

M. Ilyas et.al outlined an automated diagnosis prediction of leukemia by deeply analyzing WBC's components like Neutrophile, Eosinophile Monocytes and Lymphocytes using augmented dataset. Detection has been done by classifying normal and abnormal means blast cell. This study also compared different pre-trained model by evaluating various statistical parameters [10]

This paper [11] presented MATLAB based identification of WBC's and method correctly identification of WBCs and leucoblasts cells in the blood sample. For refining images, it uses thresholding technique and k-mean algorithm to extract features in images. This boost WBC counting and whole segmentation.

Sachin k. et.al designed an algorithm for automated image based acute leukaemia detection systems. This process utilizes image augmentation, morphology, filtering and segmenting technique to extract region of interest using k – means clustering algorithm. This Technique attained an accuracy of 92.8% and is tested with k-Nearest Neighbour and Naïve Bayes Classifier by taking 60 samples [12]

The research [13] evaluated Leukemia diagnosis with the help of statistical tool SPSS and concavity-based segmentation approaches. Significant features were then input into machine learning models, including SVM, KNN, NN, DT, and NB. The Neural Network model demonstrated superior performance, yielding an AUC of 98.9% and an accuracy of 95%.[13]

Q. Liao et.al. presented white blood cell segmentation approach that consist mathematical morphological operation, shape arrangement and detection. This system has been applied to 25 slides and 128 white blood cells segmented from background and overlapping red blood cell successfully.[14]

Vardhan et.al claimed that deep learning methods are helpful for diagnosis of blood cancer by using CNN and ANN which directly extract features of microscopic blood samples. By comparing ALL and CLL blood smear using Otsu's Thresholding method it is clear that diagnosis can be greatly enhanced.[15]

The Cytometry technique that performs automated counting fails to identify the abnormal cells. Manual detection hemacytometer are susceptible to errors and are imprecise. To overcome these challenges Vasundhara et.al [16] crafted Data mining Algorithm to detect blast cell and uses image processing for detail segmentation. The results show the classification of the acute lymphoblastic leukaemia into three categories viz.: L1, L2, L3. The model can separate a normal peripheral blood smear and an abnormal blood smear. The proposed algorithm demonstrated exceptional performance, achieving an overall accuracy of 98.6%. [16]

Dese et.al created an automated system for the detection of Leukaemia disease based on Machine learning. Their system was able to classify blood cancer into four categories with 97% accuracy [17]

Madhloom et.al [18] experimented on RGB images processed to extract the Hue (H) and Saturation (S) bands, which were then converted to binary format. A morphological opening operation was applied to the H band using fifteen disk-shaped structuring elements, while the S band underwent erosion. The images were then reconstructed using morphological operators, enabling the classification of blast cells based on shape features, specifically centroid and axis length. the system demonstrated high performance in

classifying leukemia types, achieving sensitivity, and specificity for the test datasets, and accuracy of 97.5%and for the validation datasets.[18]

Table 1: Comparison of Leukaemia Detection Methodology with Accuracy

References	Methodology	Segmentation Method	Accuracy	Feature extracted	Details
Putzu [4]	Image Processing	SVM	92%	Shape, Color and Texture	245 of 267 leucocytes detected from 33 images
Ansari [5]	CNN	CNN	99.5%	Precision, sensitivity, f1 score, accuracy	From 938 images,657 trained ,94 validited,187 for testing
Fatma [6]	Optimized CNN	Fuzzy logic	99.99%	Precision, recall, accuracy and specificity	Training set 20,000 and 10000 validations set
Ajeya bendigeri [7]	SVM	CNN	97.2%	Morphological	Not mentioned
Haque [8]	ML and TL	DCNN	96.89%	Accuracy, Precision Recall and f1 score	Inception-Res-Net model shows outstanding result.
Bodzas [9]	ANN	SVM	97.52%	Morphological and statistical	From 241 images 128 normal and 113 is infected
Llyas M [10]	DNN	CNN	99%	Accuracy, Precision Recall and fscore	From 2499 training images and 624 testing images
Christo Ananth P [11]	MATLAB	k-mean clustering	98.8%	Accuracy, Contrast stretching, Histogram equalization	Not mention
Sachin kumar [12]	Image processing	Naïve Classifier, k-mean clustering	92.8%	Statistical and morphological	60 sample images detected
Vandana et.al [13]	Neural Network	Concavity based segmentation	95%	Statistical and morphological	12 features extracted using 5 ML classifiers
Q. liao et, al [14]	Hough transformation Image processing	Thresholding	Not mentioned	Morphological features	128 white blood cells segmented

Vardhan et. al [15]	ANN	CNN	99.72%	Edges, texture, form and color	90 out of 100 samples
Vasundhara [16]	Data mining algorithm	K-means K-medoids	98.6%.	Texture, shape, and visual features are extracted.	Not mentioned
Baig [17]	Deep learning	2- CNN models	97.04%	Morphological feature	4150 images 70% data used for training and 30% for testing
Dese [18]	Image processing	CNN	97%	accuracy, sensitivity, specificity	Blood smear slides (n = 250
Madhloom [19]	Image processing	CNN	100%	Shape, color, centroid axis	180 microscopic images were tested

Discussion

Several studies detected and differentiated leucoblast from normal blood cells using machine learning algorithm and analyzed data samples by applying various segmentation techniques. Some studies focused on morphological features extraction approach, whereas other methods combined color, texture, and additional features for microscopic classification of leukemia. By using image processing techniques, effective feature mining can be achieved, potentially leading to improved classification accuracy.

Segmentation of images not only allow to study textural features but also focuses on morphological and statistical information. In this paper a complete analysis of different techniques and their accuracy has been discussed. By evaluating accuracies, it is observed that the thresholding and SVM based segmentation resulted in lowest accuracy of 92%. The proposed method involves separating leukocytes from other blood components and extract lymphocytes with segmentation methods like k-mean clustering, fuzzy clustering. It is observed that most effective method of segmentation is k-means and CNN clustering which extract appropriate features from sample data. In most of the methods accuracy range from 92%-95%. While Deep Learning algorithm namely Convolution neural network and Artificial neural network shows accuracy above 97%. It is observed that if CNN algorithm used with image processing it yields 100% accuracy of diagnosis. There is improvement area in SVM technique, further research can be used to achieve improved accuracy of SVM; while analyzing the data it is noted that to get enhanced accuracy it is necessary to take sufficient amount of data samples from authentic sources. If significant dataset accuracy may drop down. This proposed study proved that Machine Learning and Deep Learning algorithm shows maximum accuracy.

Conclusion

This paper provides a complete review of various preprocessing techniques and classification methods for leukemia detection. This paper discusses recent advancements in machine learning-based classification approaches, presenting results and analysis of existing methods. Our findings suggest that leukemia classification can be effectively achieved using machine learning algorithms. However, for significant datasets, deep learning method demonstrated superior performance is needed for making them a promising approaching towards accurate leukemia classification."

Reference

1]Huang, J., Chan, S. C., Ngai, C. H., Lok, V., Zhang, L., Lucero-Prisno, D. E., 3rd, Xu, W., Zheng, Z. J., Elcarte, E., Withers, M., & Wong, M. C. S. (2022). Disease Burden, Risk Factors, and Trends of Leukaemia: A Global Analysis. *Frontiers in oncology*, *12*, 904292. https://doi.org/10.3389/fonc.2022.904292

2] B.J. Bain, L. Estcourt, FAB Classification of Leukemia Stanley Maloy, Kelly Hughes, Brenner's Encyclopaedias of Genetics (Second Edition), Academic Press,2013, Pages 5-7, ISBN 9780080961569, https://doi.org/10.1016/B978-0-12-374984-0.00515-5.

3] RD Labati, V Piuri, F Scotti "All-IDB: The acute lymphoblastic leukemia image database for image processing" 2011 18th IEEE international conference on image processing, 2045-2048

4] Lorenzo Putzu, Giovanni Caocci, Cecilia Di Ruberto, Leucocyte classification for leukaemia detection using image processing techniques, Artificial Intelligence in Medicine, Volume 62, Issue 3,2014,

5] Ansari, S., Navin, A. H., Sangar, A. B., Gharamaleki, J. V., & Danishvar, S. (2023). A Customized Efficient Deep Learning Model for the Diagnosis of Acute Leukemia Cells Based on Lymphocyte and Monocyte Images. *Electronics*, vol.*12*(2), pp322. https://doi.org/10.3390/electronics12020322

6] Talaat, F.M., Gamel, S.A. Machine learning in detection and classification of leukemia using C-NMC_Leukemia. *Multimed Tools Appl* 83, 8063–8076 (2024). https://doi.org/10.1007/s11042-023-15923-8

7] A.Bendigeri,A. Kurabet , A.K Singh, I.Inamat , V. Jarali(2023) Comparison of Algorithms for Blood Cancer Detection Vol 4, no 12, pp 2680-2683

8] Haque, R., Al Sakib, A., Hossain, M. F., Islam, F., Ibne Aziz, F., Ahmed, M. R., Kannan, S., Rohan, A., & Hasan, M. J. (2024). Advancing Early Leukemia Diagnostics: A Comprehensive Study Incorporating Image Processing and Transfer Learning. *BioMed Informatics*, *4*(2), pp966-991.

9] Bodzas, A., Kodytek, P., & Zidek, J. (2020). Automated Detection of Acute Lymphoblastic Leukemia From Microscopic Images Based on Human Visual Perception. *Frontiers in bioengineering and biotechnology*, *8*, 1005.

10] Ilyas, M., Bilal, M., Malik, N., Khan, H. U., Ramzan, M., & Naz, A. (2024). Using Deep Learning Techniques to Enhance Blood Cell Detection in Patients with Leukemia. *Information*, *15*(12), 787. https://doi.org/10.3390/info15120787

11] Ananth, C., Tamilselvi, P., Joshy, S.A., Kumar, T.A. (2023). Blood Cancer Detection with Microscopic Images Using Machine Learning. In: Deva Sarma, H.K., Piuri, V., Pujari, A.K. (eds) Machine Learning in Information and Communication Technology. Lecture Notes in Networks and Systems, vol 498. Springer, Singapore. https://doi.org/10.1007/978-981-19-5090-2_4

12] Sachin Kumar, Sumita Mishra, Pallavi Asthana, Pragya, ―Automated Detection of Acute Leukemia Using K-mean Clustering Algorithm, Advances in Computer and Computational Sciences. Advances in Intelligent Systems and Computing, Springer vol 554, (2018) pp: 655-670.

13] Vandana Khobragade, Jagannath Nirmal, & Aayesha Hakim (2024). Statistical Analysis of Features for Detecting Leukemia. International Journal of Online and Biomedical Engineering (iJOE). 20. 130-150. 10.3991/ijoe.v20i10.47157.

14] Q. Liao & Yingying Deng, (2002). An accurate segmentation method for white blood cell images. 245 - 248. 10.1109/ISBI.2002.1029239

15]R. Vardhan, P. Poornima, M. Shrestha, Shwetha R.M., D. Soundarya (2022). Detection of ALL and CLL blood cells in WBC's using blood smear images. Vol32 Issue 1, 2023

16] Acharya, V., & Kumar, P. (2019). Detection of acute lymphoblastic leukaemia using image segmentation and data mining algorithms. *Medical & biological engineering & computing*, *57*(8), 1783–1811. https://doi.org/10.1007/s11517-019-01984-1

17] Baig, R., Rehman, A., Almuhaimeed, A., Alzahrani, A., & Rauf, H. T. (2022). Detecting Malignant Leukemia Cells Using Microscopic Blood Smear Images: A Deep Learning Approach. *Applied Sciences*, *12*(13), 6317. https://doi.org/10.3390/app12136317

18]Dese, K., Raj, H., Ayana, G., Yemane, T., Adissu, W., Krishnamoorthy, J., & Kwa, T. (2021). Accurate Machine-Learning-Based classification of Leukemia from Blood Smear Images. Clinical lymphoma, myeloma & leukemia, 21(11), e903–e914.

19] F. Scotti, "Automatic morphological analysis for acute leukemia identification in peripheral blood microscope images," *CIMSA. 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, 2005.*, Messian, Italy, 2005, pp. 96-101, doi: 10.1109/CIMSA.2005.1522835.

20]Ghaderzadeh, Mustafa, Asadi, Farkhondeh, Hosseini, Azamossadat, Bashash, Davood, Abolghasemi, Hassan, Roshanpour, Arash, Machine Learning in Detection and Classification of Leukemia Using Smear Blood Images: A Systematic Review, *Scientific Programming*, 2021, 9933481, 14 pages, 2021. **https://doi.org/10.1155/2021/9933481**